

## APPENDIX

1. Protein fold classification using patterns in protein sequences  
U: Protein sequences  
S: Sequence patterns in protein sequences.  
5 Q: Sequences known to belong to a particular fold (three dimensional arrangement in space of the amino acid chain). Example: sequences known to belong to a particular class/fold/superfamily in SCOP protein classification system. Other fold classification systems like CATH, CE, FSSP, VAST etc could also be used.  
10 Notes: While it is widely beleived that fold is a direct consequence of the sequence, successful prediction of the protein fold from the amino acid sequence remains a grand challenge in biology. Application of the method described here will provide the ability to use sequence patterns in new protein sequences to classify them into known folds.
2. Medical diagnostics using gene expression profile from DNA microarrays  
15 U: Patients, with and without disease  
S: Expression profile for each gene  
Q: Patients known to have a particular disease (clinical diagnosis).
3. Supermarket purchases  
U: Purchases (a purchase is the contents of the cart)  
20 S: Each item in the store  
Q: Time of day/ location/ age group/ mode of payment
4. Airline flights  
U: Commercial airline flights  
S: Airline  
25 Departure city  
Arrival city

	Scheduled time of day/day of week/month/year of departure
	Scheduled time of day/day of week/month/year of arrival
	Crew composition
	Equipment
5	Weather at departure/arrival/enroute
	Occupancy
	Scheduled layover from previous flight
	Scheduled flying time
	Q: On time departure
10	On time arrival
	Safety incidents
	Customer satisfaction
	5. Fast track security check at airports for frequent travellers
	U: Travelling individuals
15	S: Flight attributes as in Application 4.
	Individual attributes:
	Age
	Sex
	Height
20	Weight
	Name
	Checked in bags
	Carry on bags
	Price of ticket
25	Class of travel
	Mode of payment

- Mode of purchase (travel agent/on line/airline)  
Advance purchase  
Accompanying passengers  
Connecting to/from other flights
- 5 Q: Found violating security requirements
6. Automobile insurance
- U: Automobile insurance policies
- S: Auto: make, model, year, trim, price paid, color, condition, bought/leased  
Geography: residence zip code, work zip code, commute distance
- 10 Driver: Individual characteristics as in application 5  
Years of driving experience  
Previous accidents - caused, involved(not caused)  
Points in driver's license  
insurance claim history/record
- 15 Q: Makes insurance claim  
collision  
damage  
theft  
personal injury
- 20 7. Matchmaking (example: dating service)
- U: Match events, each constituting of a pair of individuals
- S: Individual characteristics
- Q: Satisfactory match

8. Matching of buyer-seller in on-line trading like eBay  
U: Trading events  
S: Individual characteristics  
Traded object characteristics
- 5 Q: Difference between asking and traded price  
Satisfaction with trade  
Satisfaction with traded object
9. Advertisements on Web Portals  
U: web pages
- 10 S: content keywords/indices  
Browser( domain, time of day, comes from)  
Q: likelihood of clicking on particular/class of ad  
viewing the page
10. Protein function classification U: Protein sequences
- 15 S: Sequence patterns or other attributes in protein sequences.  
Q: Sequences known to belong to a particular functional fold.
11. Product Placement
12. Health/personal product selection  
U: patient/consumer, product/service
- 20 S: Genotype/lifestyle of patient/consumer  
Q: Product/service

13. Identification of communication intercepts that constitute security threat
- U: Communication intercepts
- S: keywords, patterns, triggers
- origin, destination of communication
- 5 medium (cell phone/land phone/email/mail)
- time, day, date
- Q: known to constitute security threat
- pertain to some monitored activity (surveillance of known terrorist/drug trafficking networks)
- 10 14. Will a new patient respond well to known treatment?
- Design/improve clinical trials by identifying and populating new patients into the non-overlapping set(s) that cause the most error.
- U: Clinical patients
- S: Symptoms
- 15 Individual characteristics (age, sex, location, lifestyle, etc)
- Q: Patients known to respond to a particular treatment
15. Identifying the anonymous/misattributed authors of texts
- U: Complete texts (example books, poetry)
- S: patterns of words (example discovered by pattern discovery)
- 20 Q: Texts known to be by a particular author
- of a particular genre/period
- Example of relevance of this application in
- <http://w3.research.ibm.com/visions/foster/foster.html>